Aplicación de Hojas Electrónicas para el cálculo de las frecuencias

Dr. Servio Interiano, Octubre 2021.

Introducción

La aplicación de las Hojas Electrónicas para el trabajo de series simples de datos en estadística, es una excelente ayuda cuando se trabajan con una cantidad relativamente moderada de datos. Para datos voluminosos (miles de datos), los programas de análisis estadísticos como R son imprescindibles. Pero para trabajos de investigación universitarios una Hoja Electrónica, como Excel de Microsoft Office, Calc de Libreoffice, Google Sheets o Gnumeric es suficiente.

El cálculo de las estadísticas descriptivas es muy sencillo con las funciones que incluyen la mayoría de Hojas Electrónicas, sin embargo el elaborar distribuciones de frecuencias puede ser un poco laborioso si estamos trabajando con un gran número de variables.

Afortunadamente existe una forma sencilla de calcular las frecuencias de una serie de datos simples, sin tener que pasar por el proceso "manual" de ordenar los datos y contar los que caen en cada intervalo de clase.

En el presente documento voy a presentar el proceso para la elaboración de Distribuciones de Frecuencias para series simples de variables cuantitativas continuas, con cada una de las Hojas Electrónicas mencionadas.

El orden de la presentación será iniciando con Gnumeric por ser una Hoja Electrónica de Software Libre multiplataforma que puede descargarse tanto en Windows como en Linux, luego Libreoffice Calc, Google Sheet y por último Microsoft Excel.

Índice

Introducción **1**

Índice **2**

Descripción General **2**

Gnumeric **3**

Libreoffice Calc **8**

Google Sheets **12**

Microsoft Excel **15**

Descripción General

Al emplear Hojas Electrónicas para el cálculo de las Estadísticas Descriptivas, contamos con funciones que nos facilitan los cálculos, de esta manera para el cálculo de la media aritmética sólo debemos escribir en la celda donde deseamos el resultado la función “=average(rango)”, donde rango significa marcar el rango de datos de los datos que queremos analizar.

Otras funciones para la obtención de datos claves son: min, max, count, median, mode, skew, kurtosis, stdev, var, etc.

Junto con éstas funciones encontramos una función muy poderosa y no tan conocida para obtener fácilmente la frecuencia con la que aparece un dato dentro de una lista y agruparlo dentro de una categoría. Esto es conocido en estadística como distribución de frecuencias.
Una distribución de frecuencias no es más que un conteo de datos para saber cuáles de ellos caen dentro de las categorías o intervalos de clase.

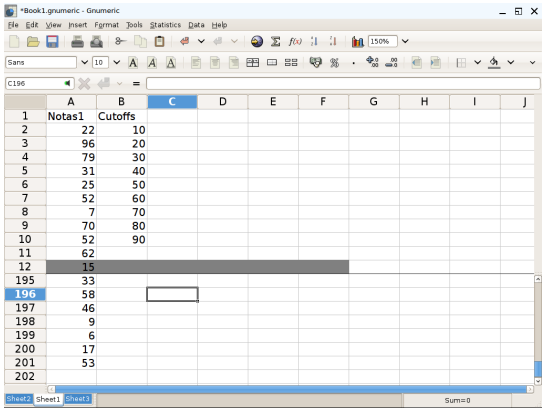
En la Hoja Electrónica Gnumeric tenemos un módulo de análisis que nos realiza el cálculo de las frecuencias para una serie de intervalos, lo cual se logra con la selección de los rangos y rompimientos (cutoff) de los intervalos a través de un menú de opciones.
En el resto de las Hojas Electrónicas debemos “programar” la función en una celda con un proceso específico, pero que no es realmente nada muy complicado. Explicaremos el proceso en cada una de las Hojas Electrónicas seleccionadas.

Gnumeric

Según encontramos en la Wikipedia: “Gnumeric es una hoja de cálculo libre que forma parte del entorno de escritorio libre GNOME. La versión 1.0 de Gnumeric fue publicada el 31 de diciembre de 2001. Gnumeric es distribuida como software libre bajo la licencia GNU GPL; Su intención es, junto con todo el software libre, reemplazar al software privativo y hojas de cálculo no libres como Microsoft Excel.”

Para calcular nuestras frecuencias con Gnumeric, tomaremos una lista de calificaciones de alumnos que pueden ir desde 1 hasta 100. Ahora queremos saber cuántos alumnos obtuvieron una calificación entre 0 y 10, y también cuántos alumnos obtuvieron una calificación entre 11 y 20 y así sucesivamente hasta llegar a 100.

En este ejemplo, en la columna A tenemos la lista de calificaciones de 200 alumnos, a la que llamaremos “Notas1” y en la columna B los rangos en los que deseamos agrupar dichas calificaciones, a los que llamaremos “cutoffs”.



	A	B
1	Notas1	Cutoffs
2	22	10
3	96	20
4	79	30
5	31	40
6	25	50
7	52	60
8	7	70
9	70	80
10	52	90
11	62	
12	15	
195	33	
196	58	
197	46	
198	9	
199	6	
200	17	
201	53	
202		

Figura 01

Vemos en la Figura 01, la pantalla de trabajo de Gnumeric, en donde tenemos ya nuestros datos en la columna A. Observamos los primeros 12 datos, luego un corte y los últimos 6 datos.

En la columna B, observamos en las primeras 10 filas los Límites Aparentes Inferiores de cada intervalo, donde queremos hacer los cortes de los intervalos (cutoffs).

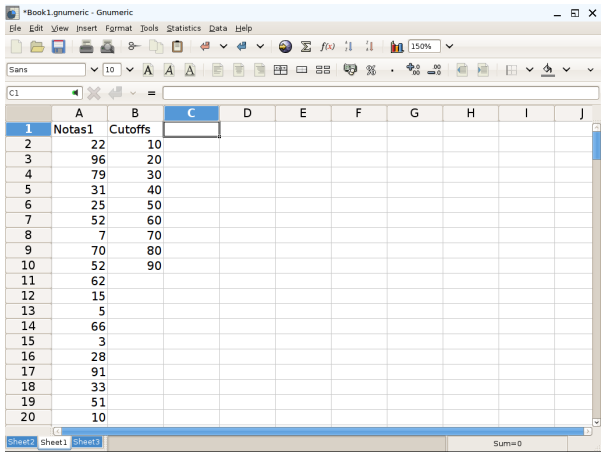


Figura 02

Vamos a trabajar con la imagen de la Figura 02, donde se aprecian únicamente los primeros datos, con la salvedad que se están incluyendo la totalidad de los datos.

Lo primero que vamos a hacer es seleccionar el módulo para el cálculo de las frecuencias. Esto lo hacemos seleccionando en el menú:

- Statistics
- Descriptive Statistics
- Frequency tables
- Histogram

Como se muestra en la Figura 03.

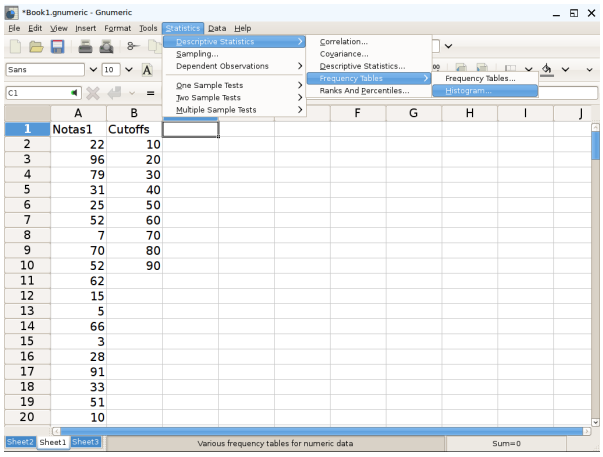


Figura 03

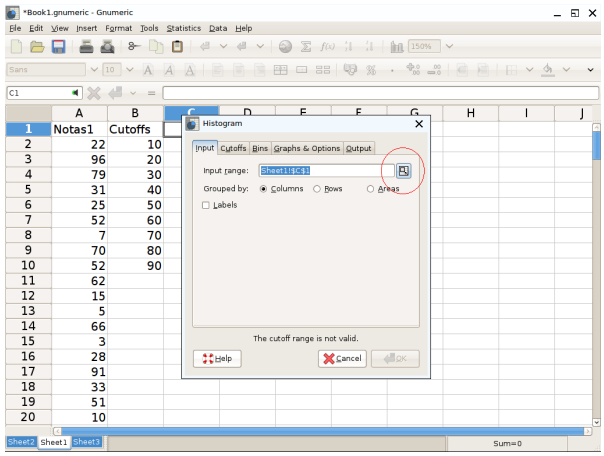


Figura 04

Esto nos abrirá una ventana flotante, en donde controlaremos el rango de datos a ingresar, los intervalos a utilizar y el lugar donde colocaremos nuestras frecuencias resultantes.

Empezaremos por seleccionar los datos a incluir.

Seleccionamos el ícono para la selección del rango de datos.

Este se muestra encerrado en un círculo rojo en la Figura 04.

Seguidamente nos regresará a la Hoja de Trabajo de Gnumeric, en donde podemos seleccionar la serie de datos a incluir.

Al finalizar hacemos “click” en el ícono de selección de datos de la ventana flotante remanente, Figura 05

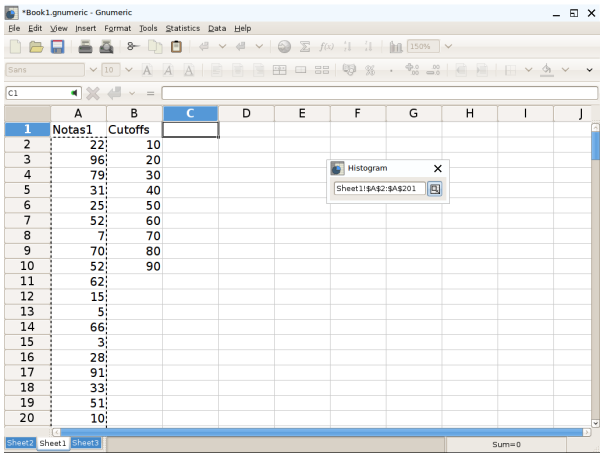


Figura 05

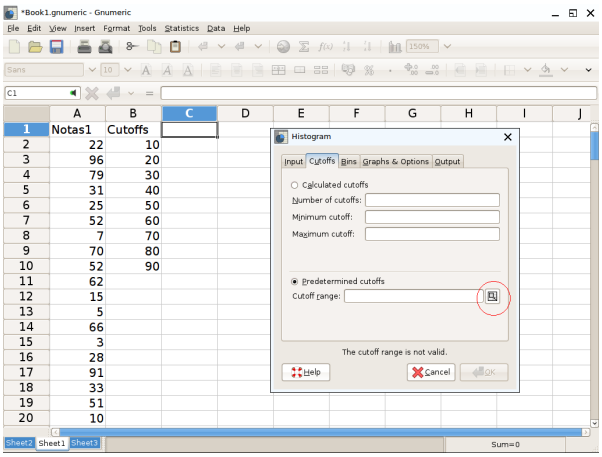


Figura 06

Regresamos a ventana flotante y seleccionamos la segunda pestaña, en donde hacemos “Click” en el ícono para la selección del corte de los intervalos de clase (cutoffs).

Se muestra encerrado en un círculo rojo en la Figura 06.

Una vez más nos devuelve a la Hoja de Trabajo de Gnumeric, en donde seleccionamos los Límites Aparentes inferiores que habíamos colocado en la columna B.

Después de señalarlos, volvemos a la ventana flotante haciendo “Click” en el ícono de selección de los datos.

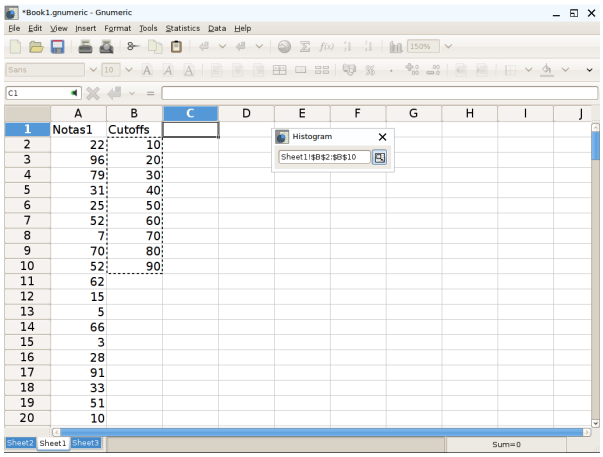


Figura 07

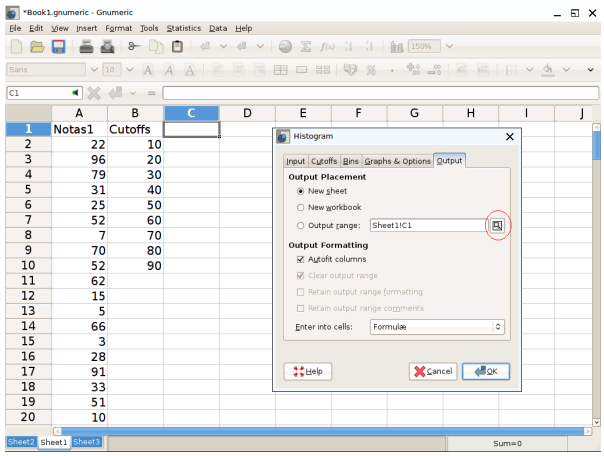


Figura 08

Al regresar a la ventana flotante, seleccionamos la última pestaña, para determinar en donde queremos nuestros intervalos y sus frecuencias.

Por omisión el programa los coloca en una hoja nueva, vamos a hacer “Click” en el ícono para la selección de la celda de salida para cambiar el valor por omisión.

Se muestra encerrado en un círculo rojo en la Figura 08.

Seleccionamos la celda de inicio para la colocación de los resultados del cálculo de las frecuencias, señalando con el mouse y haciendo “Click” en el ícono de selección de la ventana flotante.
Figura 09

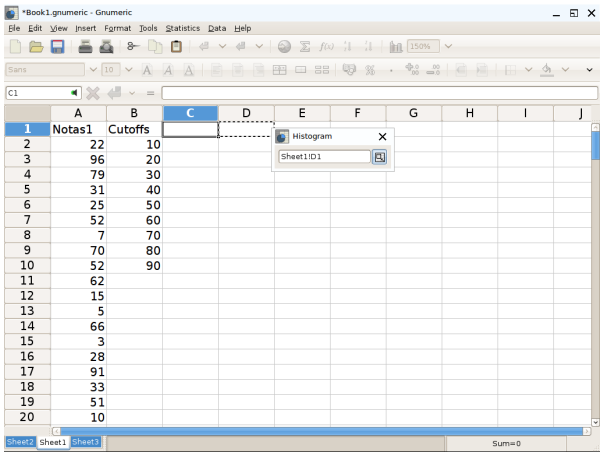


Figura 09

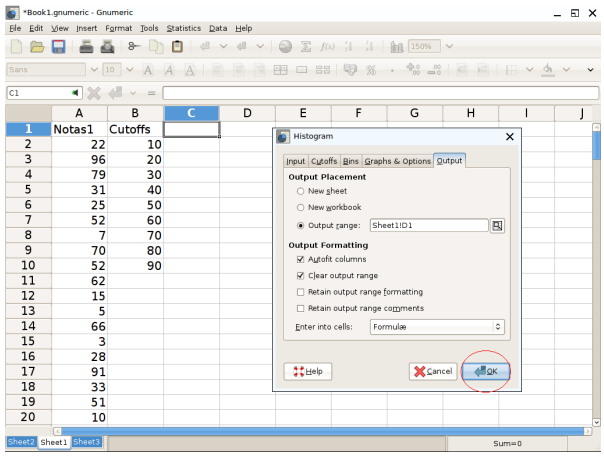
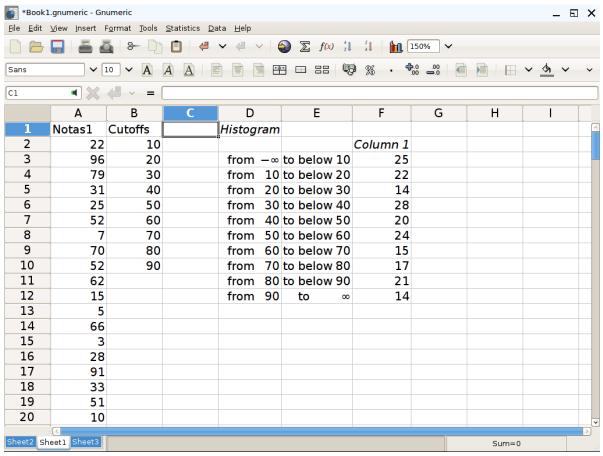


Figura 10

Por último hacemos “Click” en el botón “OK” para finalizar el proceso y calcular las frecuencias de los datos en cada uno de los intervalos de clase. Figura 10

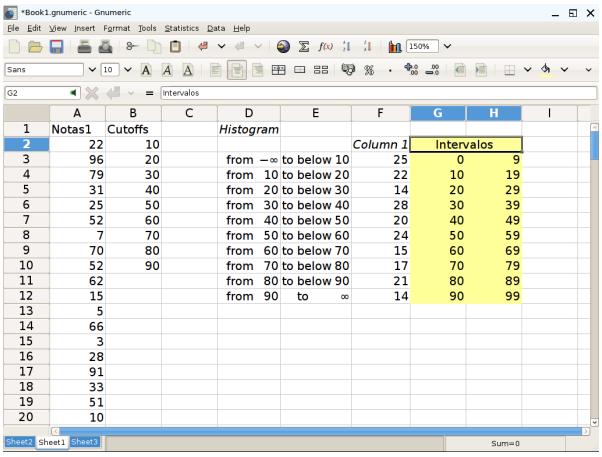
En la Figura 11 podemos apreciar que el programa incluye tres columnas:

- La primera con el límite inferior del intervalo. “from x”
- La segunda con el límite superior del intervalo. “to below x”
- La tercera con las frecuencias correspondientes a cada intervalo de clase.



	A	B	C	D	E	F
1	Notas1	Cutoffs		Histogram		
2	22	10				Column 1
3	96	20		from −∞	to below 10	25
4	79	30		from 10	to below 20	22
5	31	40		from 20	to below 30	14
6	25	50		from 30	to below 40	28
7	52	60		from 40	to below 50	20
8	7	70		from 50	to below 60	24
9	70	80		from 60	to below 70	15
10	52	90		from 70	to below 80	17
11	62			from 80	to below 90	21
12	15			from 90	to ∞	14
13	5					
14	66					
15	3					
16	28					
17	91					
18	33					
19	51					
20	10					

Figura 11



	A	B	C	D	E	F	G	H
1	Notas1	Cutoffs		Histogram				
2	22	10				Column 1	Intervalos	
3	96	20		from −∞	to below 10	25	0	9
4	79	30		from 10	to below 20	22	10	19
5	31	40		from 20	to below 30	14	20	29
6	25	50		from 30	to below 40	28	30	39
7	52	60		from 40	to below 50	20	40	49
8	7	70		from 50	to below 60	24	50	59
9	70	80		from 60	to below 70	15	60	69
10	52	90		from 70	to below 80	17	70	79
11	62			from 80	to below 90	21	80	89
12	15			from 90	to ∞	14	90	99
13	5							
14	66							
15	3							
16	28							
17	91							
18	33							
19	51							
20	10							

Figura 12

Éstas columnas las podemos traducir como se muestran en las columnas de color amarillo de la Figura 12.

“from -∞ to below 10” significa menor que 10, lo que para nuestros datos significa de 0 a 9.

“from 10 to below 20” significa de 10 hasta por debajo de 20, que para nuestros datos serían de 10 a 19. Etc.

Y listo!

Tenemos nuestras frecuencias calculadas sin necesidad de ordenar los datos o realizar el conteo manualmente entre los datos contenidos en cada intervalo.

Libreoffice Calc

Según se describe en Wikipedia: “LibreOffice Calc es la hoja de cálculo del paquete de software LibreOffice y es una bifurcación de OpenOffice.org Calc. Es una hoja de cálculo similar a Microsoft Excel y a Quattro Pro de Corel.

Después de bifurcarse de Openoffice.org en 2010, Calc experimentó una masiva revisión para corregir un número grande de defectos en cálculos de fórmulas que implicaban referencias externas, y para aumentar el desempeño de cacheo de datos, especialmente cuando se referenciaban rangos de datos grandes.

Adicionalmente, ahora Calc soporta 1 millón de filas en una hoja de cálculo con referencias de macro en cada celda.

Calc es capaz de abrir y guardar la mayoría de las hojas de cálculo en el formato de archivo de Microsoft Excel. También es capaz de guardar las hojas de cálculo como archivos PDF.

Como con toda la suite LibreOffice, Calc está disponible para una variedad de plataformas, incluyendo Linux, OS X, Microsoft Windows y FreeBSD, bajo la Licencia Pública General Menor GNU. Calc es un software libre y de código abierto.”

A diferencia de Gnumeric, Calc no tiene un módulo dedicado al cálculo de las frecuencias, pero posee una función que realiza la misma función con algunas pequeñas diferencias.

La función para el programa en español es:

=Frecuencia(Rango de datos, Rango de intervalos)

Para el programa en inglés es:

=Frequency(Rango de datos, Rango de intervalos)

Como es una operación matricial, al finalizar de ingresarla debemos oprimir:

[Shift]+[Ctrl]+[Enter]

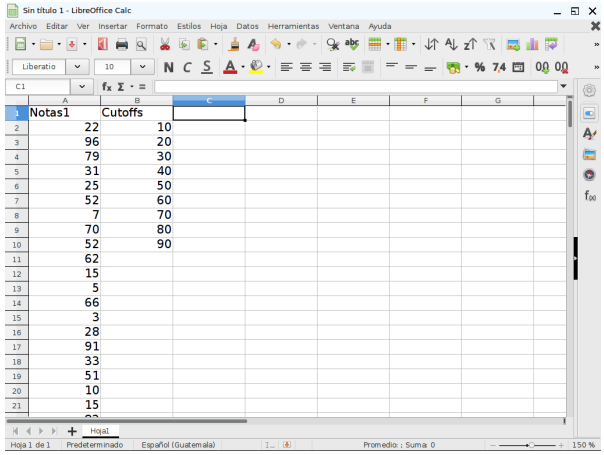


Figura 13

Vamos a iniciar con los mismos datos que teníamos en Gnumeric.

En la columna A colocamos nuestros datos, que en este caso serán notas de 200 alumnos entre 1 y 99.

En la columna B, tenemos los Límites aparentes inferiores de los intervalos que vamos a utilizar.

Antes de iniciar ...

LibreOffice Calc, a diferencia de Gnumeric no utiliza los Límites Aparentes inferiores de los intervalos de clase, sino que los Límites Reales superiores de los intervalos de clase.

Por lo que debemos modificar nuestros “cutoffs” y colocar los límites reales superiores. Por lo que debemos agregar el último que en nuestro caso será 99.5. Los datos deben quedar como se muestra en la Figura 14.

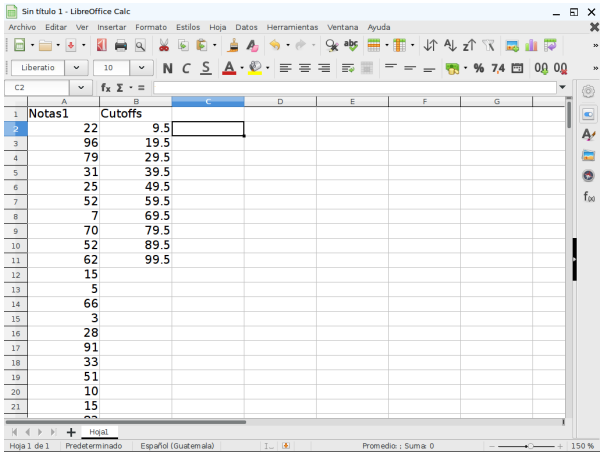


Figura 14

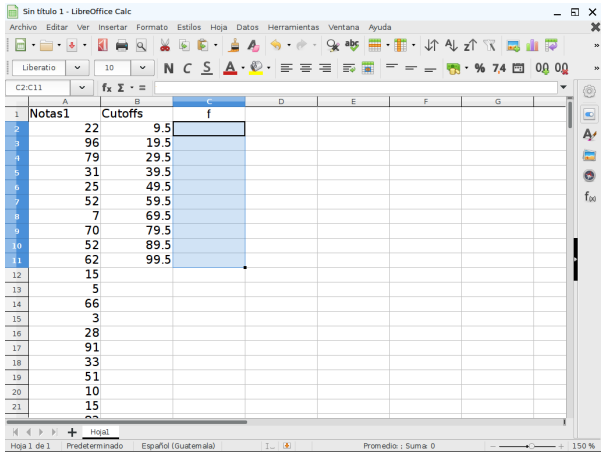


Figura 15

Lo primero que debemos hacer es marcar las celdas donde queremos que se coloquen nuestros intervalos.
Si no las marcamos tampoco pasa nada grave, el programa colocará una frecuencia “0” después el último Límite Real.
Sin embargo, para evitar esa frecuencia 0 que pudiera crear confusión, es conveniente marcarlos primero.

Una vez con el cursor en la primera fila y con las demás filas marcadas para la salida de nuestras frecuencias, empezamos a escribir nuestro comando:

=frecuencia(

Luego marcamos el rango de datos y escribimos una coma “,”

Luego marcamos los límites reales superiores de nuestros intervalos, como se muestra en la Figura 16

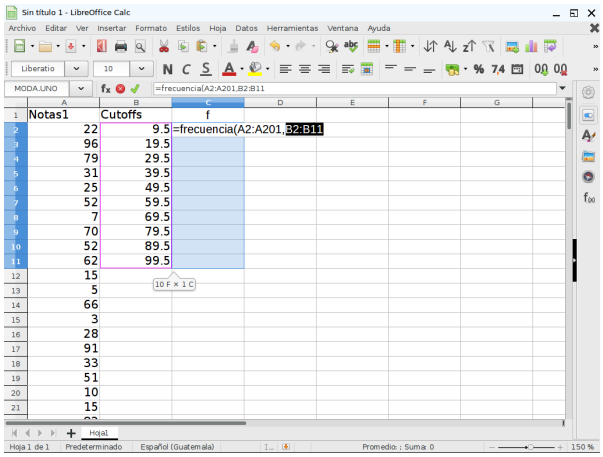


Figura 16

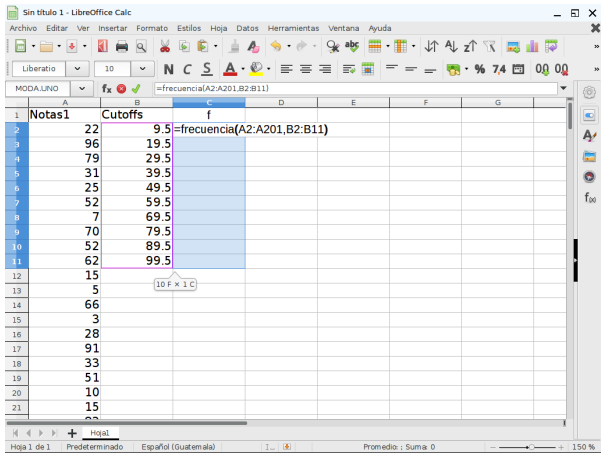


Figura 17

Terminamos de programar nuestra función cerrando el paréntesis como se muestra en la Figura 17.

El paso siguiente es oprimir
[Shift]+[Ctrl]+[Enter]

Esto es al mismo tiempo la tecla de mayúsculas, la tecla control y la tecla enter y no solo la tecla [Enter] para que tome el comando matricial.

Al oprimir [Shift]+[Ctrl]+[Enter], las celdas que habíamos marcado con anterioridad se llenan con las frecuencias correspondientes a cada intervalo de clase.

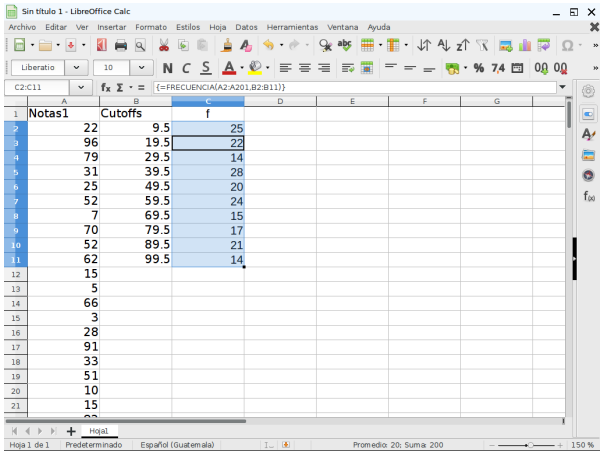


Figura 18

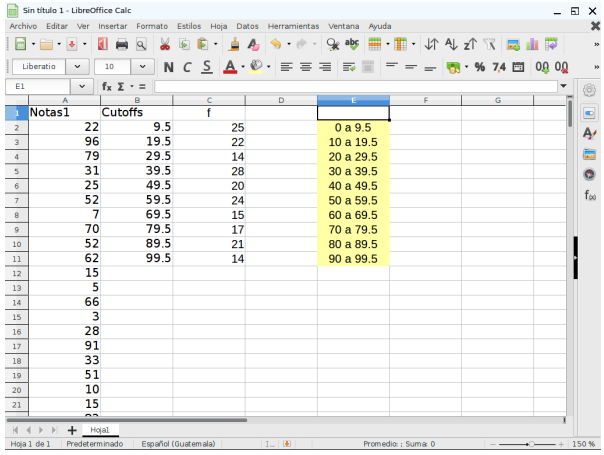


Figura 19

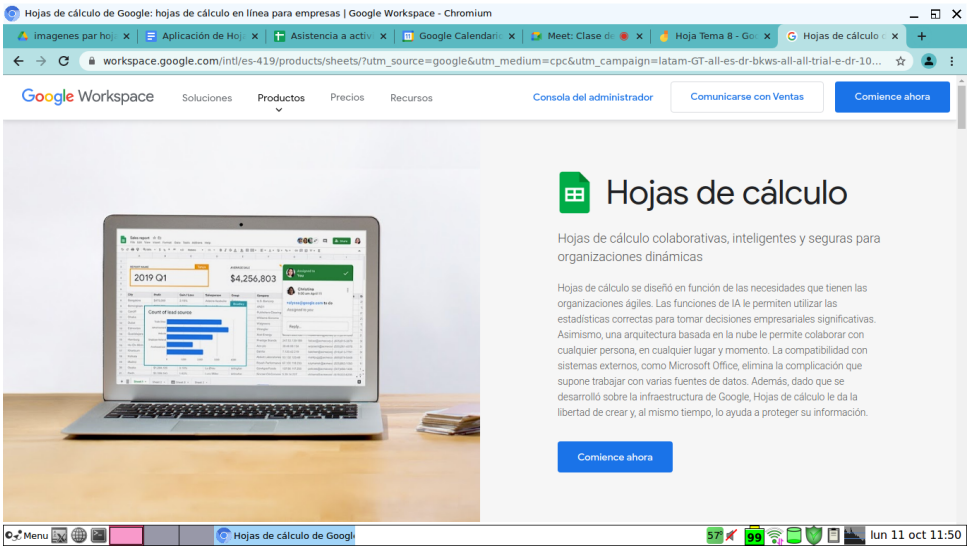
En la Figura 19, en la columna de color amarillo se muestra la traducción de nuestros intervalos.
El primero abarcaría todos los datos por debajo de 9.5, es decir de 0 a 9.

El segundo abarcaría los datos superiores a 9.5 y menores a 19.5, es decir de 10 a 19, etc.

Y listo!

Tenemos nuestras frecuencias calculadas sin necesidad de ordenar los datos o realizar el conteo manualmente entre los datos contenidos en cada intervalo.

Google Sheets



Google Sheets es un programa de hoja de cálculo incluido como parte del paquete gratuito de editores de documentos de Google basado en la web que ofrece Google.

La aplicación es compatible con los formatos de archivo de Microsoft Excel .

La aplicación permite a los usuarios crear y editar archivos en línea mientras colaboran con otros usuarios en tiempo real. El usuario realiza un seguimiento de las ediciones con un historial de revisión que presenta los cambios. La posición de un editor se resalta con un color y un cursor específicos del editor y un sistema de permisos regula lo que pueden hacer los usuarios. Las actualizaciones han introducido funciones que utilizan el aprendizaje automático , incluido "Explorar", que ofrece respuestas basadas en preguntas de lenguaje natural en una hoja de cálculo.

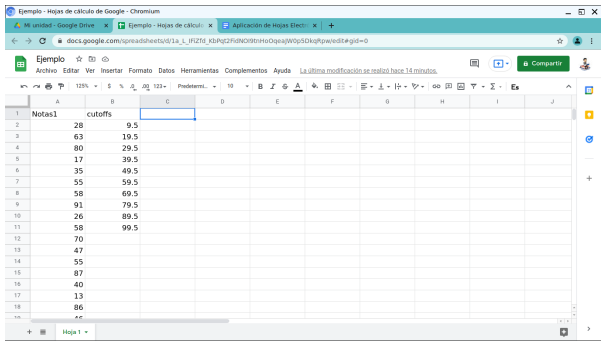


Figura 20

Vamos a iniciar igual que en los ejemplos anteriores con una hoja y dos columnas, en la columna A tenemos las notas de 200 alumnos del 1 al 99.

En la columna B, los Límites Reales Superiores de los intervalos. Google Sheets opera los datos como el LibreOffice Calc.

Lo primero que vamos a hacer es marcar el rango de datos donde queremos la salida de nuestras frecuencias.

Esto nos facilitará la programación de la función.

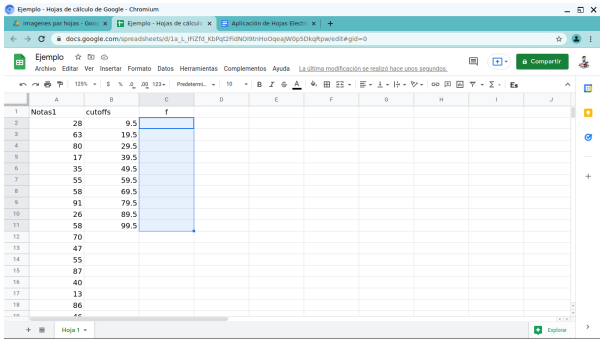


Figura 21

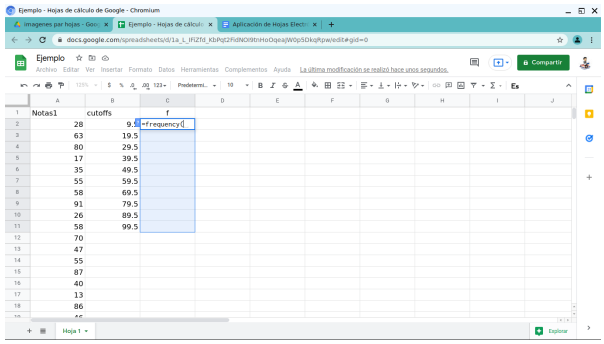


Figura 22

Luego en la primera celda de las que tenemos marcadas empezamos a ingresar nuestra función:

=Frequency(

Luego marcamos el rango donde se ubican los datos que vamos a analizar.

Marcamos los datos que vamos a analizar y escribimos una coma “,” y luego marcamos los Límites Reales Superiores de nuestros intervalos y cerramos nuestra función cerrando el paréntesis.

Luego oprimimos:
[Shift]+[Ctrl]+[Enter]

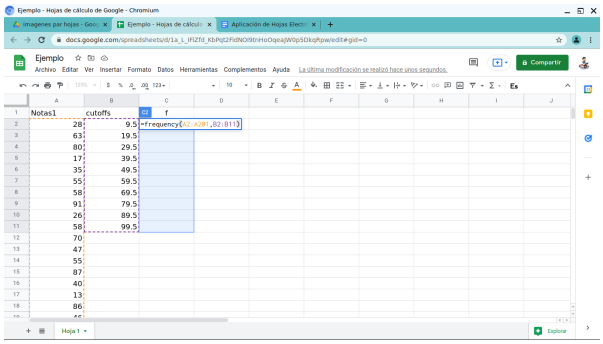


Figura 23

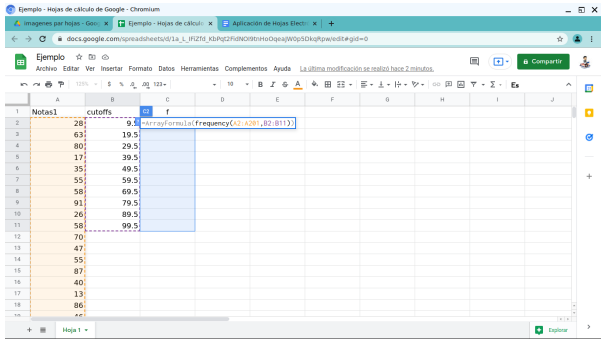


Figura 24

Al oprimir [Shift]+[Ctrl]+[Enter], el programa activa el cálculo matricial y para operarlo sólo debemos oprimir un último [Enter]

En la Figura 25 tenemos ya las frecuencias calculadas.

Nótese que el programa al final coloca una frecuencia 0. Simplemente ignoramos esta última frecuencia.

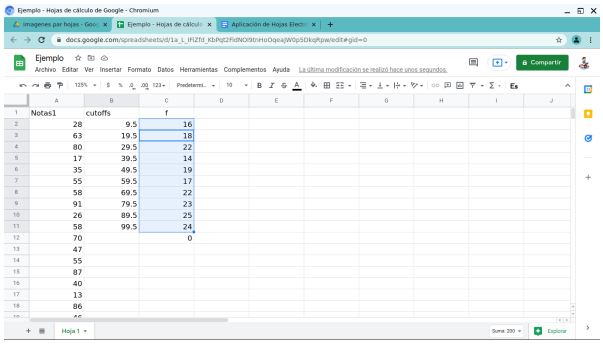


Figura 25

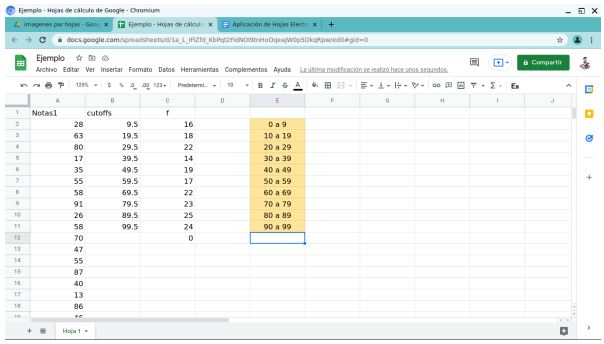


Figura 26

Al igual que con el programa LibreOffice Calc, los intervalos los podemos “traducir” como se muestran en la columna de color amarillo de la Figura 26.

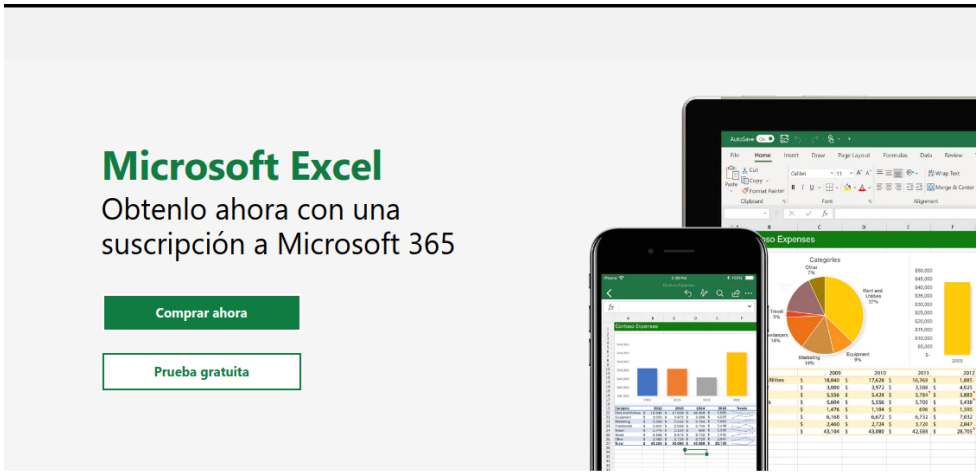
El límite real superior de 9.5 significa que contiene todos los datos de 0 a 9.

El límite real superior de 19.5 significa que contiene todos los datos de 10 a 19. Etc.

Y listo!

Tenemos nuestras frecuencias calculadas sin necesidad de ordenar los datos o realizar el conteo manualmente entre los datos contenidos en cada intervalo.

Microsoft Excel

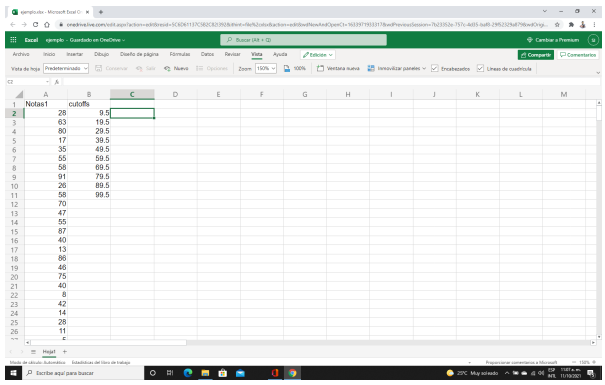


Según Wikipedia: “Microsoft Excel es una hoja de cálculo desarrollada por Microsoft para Windows, macOS, Android e iOS.

Cuenta con cálculo, gráficas, tablas calculares y un lenguaje de programación macro llamado Visual Basic para aplicaciones. Ha sido una hoja de cálculo muy aplicada para estas plataformas, especialmente desde la versión 5 en 1993, y ha reemplazado a Lotus 1-2-3 como el estándar de la industria para las hojas de cálculo. Excel forma parte de la suite de software Microsoft Office.”

Con Excel también podemos calcular las frecuencias de los intervalos aplicando la función Frequency o Frecuencia, dependiendo de si nuestro software está en inglés o español y prácticamente se opera igual que lo visto anteriormente con LibreOffice Calc y Google Sheets.

Dependiendo de la versión del programa podría ser necesario trabajarlo como lo descrito anteriormente para el LibreOffice Calc



Notas1	cutoffs
28	9.5
63	19.5
80	29.5
17	39.5
35	49.5
55	59.5
58	69.5
91	79.5
26	89.5
58	99.5
70	
47	
55	
87	
40	
13	
86	
46	
75	
40	
8	
42	
14	
28	
11	

Figura 27

Vamos a iniciar igual que en los ejemplos anteriores con una Hoja Electrónica, en la cual en la columna A tenemos las notas de 200 alumnos del 1 al 99.

Y en la columna B los Límites Reales Superiores de los intervalos de clase. Al igual que se hace con LibreOffice Calc o Google Sheets.

En este caso no es necesario marcar previamente las celdas donde vamos a querer los resultados del cálculo de las frecuencias. (En algunas versiones es necesario marcarlo como lo descrito para el LibreOffice Calc)

En la primera celda donde queremos los resultados empezamos a escribir la función:

=Frecuencia(
Y marcamos los datos.

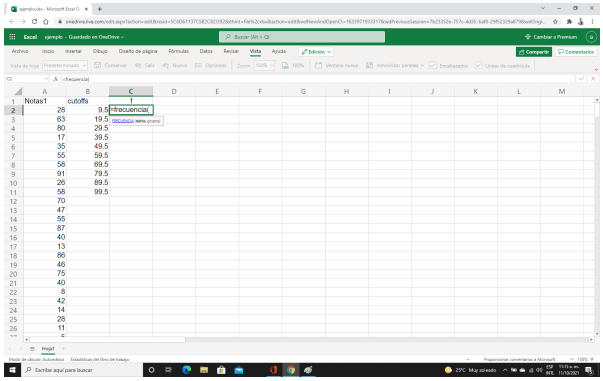


Figura 28

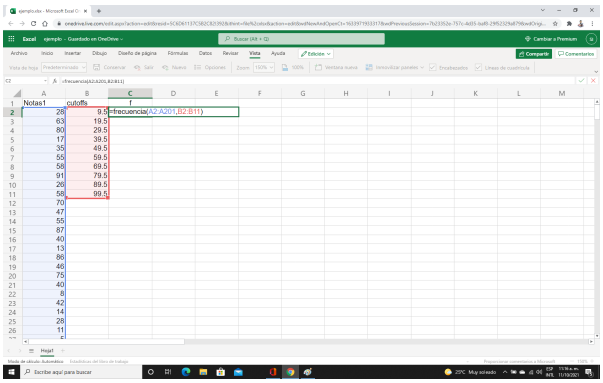


Figura 29

Luego escribimos un punto y coma “;” y marcamos los Límites Reales Superiores de los intervalos de clase.

Cerramos el paréntesis y oprimimos [Enter]

Dependiendo de la versión de Excel puede ser necesario al igual que con LibreOffice Calc y Google Sheets oprimir [Shift]+[Ctrl]+[Enter] , en las versiones más nuevas de Excel sólo es necesario oprimir [Enter]

Y obtenemos los cálculos de las frecuencias, como se muestran en Figuras 30.

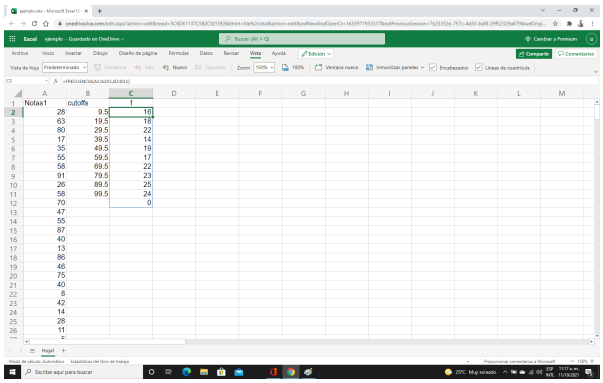


Figura 30

La interpretación de los intervalos de clase es igual que con LibreOffice Calc o Google Sheets Figuras 19 y 26.

Y listo!

Tenemos nuestras frecuencias calculadas sin necesidad de ordenar los datos o realizar el conteo manualmente entre los datos contenidos en cada intervalo.